Proceedings, 16th International ISSAT on Reliability and Quality in Design, 5-7 August,2010

Applications of Data Fusion in Monitoring Inaccessible Ocean Machinery

Janell Duhaney, Taghi M. Khoshgoftaar, Ionut Cardei, Bassem Alhalabi, John C. Sloan
Florida Atlantic University, Boca Raton, Florida, USA
Contact author: khoshgof@fau.edu

Keywords: sensor data fusion, machine condition monitoring.

***Abstract*—Machine condition monitoring (MCM) enables real time health assessment, prognostics, and advisory generation by interpreting data from sensors installed on the machine being monitored. To effectively utilize measurements for determining the health of individual components, macro-components and the overall system, these measurements must somehow be combined or integrated to form a holistic picture. The process used to perform this combination is called sensor data fusion. While research on data fusion techniques spans across many domains, not much work has been done in data fusion for MCM in unmanned systems. This paper addresses this need by providing the intuition behind a mathematical model and process for data fusion in ocean machinery.**

I. Introduction

FINDING alternative fuel sources is a worldwide initiative. One such alternative involves harvesting the natural energy created by ocean currents using ocean turbines. Efficient machine condition monitoring (MCM) systems for predictive maintenance and prognostic health monitoring (PHM) systems for predicting the future health of these ocean systems are needed to ensure system reliability and reduce operating costs. An MCM system enables real time health assessment, prognostics, and advisory generation by continuously recording and processing streams of measurements taken from sensors attached to components of the machine being monitored.The independent data from these sensors must then be combined to determine the state of individual components and, on a higher level, to determine the health of the overall system. The process of combining data from multiple sources (whether disparate or similar) to provide a more accurate and holistic solution or view is called data fusion. To implement a complete MCM/PHM system, a combination of data fusion, feature extraction, classification and prediction algorithms are needed.

Research on data fusion techniques, architectures, and approaches spans across many domains including medical diagnosis [2], military defense and tracking systems [3], robotics, navigation systems for autonomous vehicles [4], and remote sensing, but not much work, to the author's knowledge, has been done in data fusion for MCM/PHM in unattended, inaccessible[1] systems such as ocean turbines. Such systems have a low tolerance to false positives (or a false alarm) because of the expenses associated with retrieval of the equipment. The majority of research effort invested in data fusion systems for autonomous systems (whether on land or sea) were related to designing navigational systems for these vehicles. This paper is, therefore, not only unique in the target system (ocean turbines) but also in its approach to providing intuition behind a mathematical data fusion model for MCM/PHM systems for autonomous oceanic systems. This mathematical model will be formalized in future work.

This paper will attempt to formalize a data fusion approach to MCM/PHM systems for ocean turbines. In Section II, we will analyze the intended system architecture and review related work. Section III will attempt to bring existing techniques and approaches together to form a data fusion model which we will apply to our case study in Section IV. Finally, in Section V, we will highlight opportunities for future work.

II. Background/Related Work

To determine the condition of the machine, the mechanical vibration levels, oil debris, temperature, pressure, angular velocity, and flow around the ocean turbine need to be constantly monitored and analyzed. While most mechanical defects can be determined through analysis of the vibration data generated from sensors known as accelerometers, the data gathered from the remaining sensors are necessary to fully assess the state, life expectancy, and potential failure of the machine [5].

This paper assumes a system architecture satisfying the Open Systems Architecture for Condition Based Monitoring (OSA-CBM) specification. The OSA-CBM standard[2] implements the ISO-13374 standard, which defines the functionality of a condition monitoring system in terms of six blocks, to include data structure and interface method specifications for the ISO-13374 blocks [6]. These blocks are:

1) Data Acquisition (DA) – Data is collected and digitized.
2) Data Manipulation (DM) – Typically, in the DM block, signal processing (using techniques like Fast Fourier Transforms, filtering, and windowing), time synchronous averaging (TSA), algorithmic computations, and feature extraction are performed on the digitized output from the DA block.
3) State Detection (SD) – The output from DM and DA are compared against the anticipated baseline profile values

[1]With a goal of a one year maintenance free operation with annual preventative maintenance

[2]Available for download on the Machinery Information Management Open Systems Alliance (MIMOSA) website - http:\\www.mimosa.org\

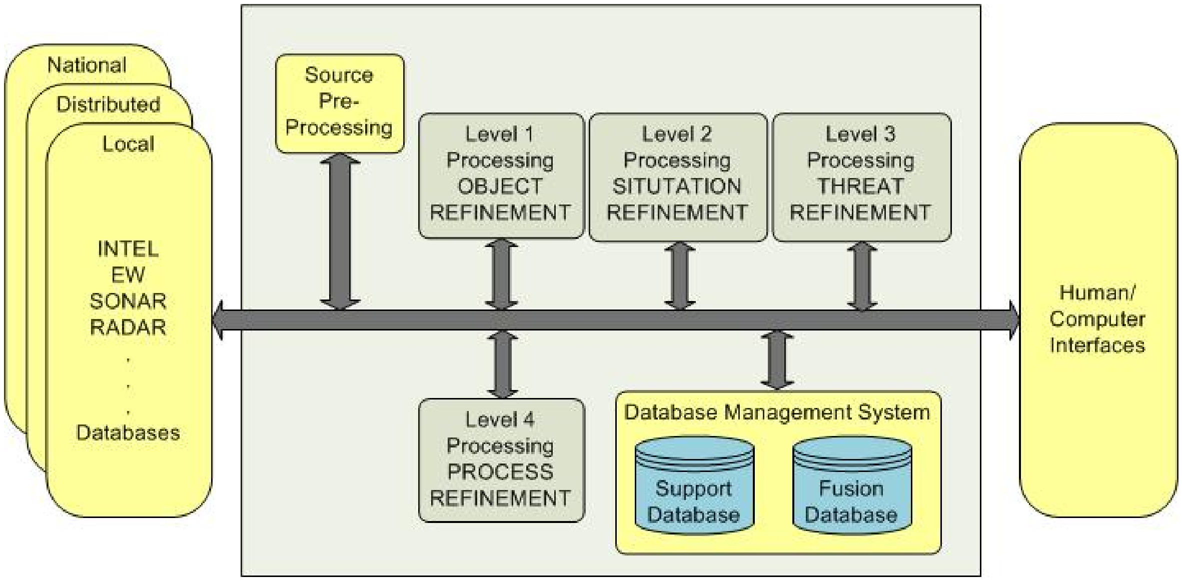

Figure 1. JDL Process Model for Data Fusion

to evaluate the system's state in terms of a predefined enumeration, e.g., system normal, level high, alarm, alert, etc.

4) Health Assessment (HA) – Diagnosing system faults and determining the health of the system occurs in this block. The output from the DA, DM and SD blocks are fused with the output from other HA blocks in order to make this assessment.
5) Prognostics Assessment (PA) – The life expectancy and future health of the system are projected, and the possibility of future faults and failures is predicted.
6) Advisory Generation (AG) – Reports on existing or predicted conditions along with advice on how to optimize the life of the machine are generated.

A. Data Fusion

Data Fusion is considered a cross cutting concern of the above architecture because the data being integrated can be provided by several entities and layers in the CBM architecture [7]. To streamline and standardize the design and codification of data fusion systems, experts in the area have proposed more than 30 fusion architectures over the years. Some of these models are discussed in [8]. Of these, one of the most widely cited model is that of the American Joint Directors of Laboratories, or JDL [3], which was initially developed for military applications. The JDL model [9] divides the processes, functions, and techniques applicable to data fusion into five levels (as seen in Figure 1). These are:

Level 0. Pre-Processing - A level 0 data preparation /estimation process would estimate entity features from one or more entity signal observations [10].

Level 1. Object Assessment - In this phase, an attempt will be made to locate and/or identify the object of interest by fusing information about this object that was gathered from multiple sources. The object assessment level is itself broken down into four substeps, namely:

- i Data Alignment - Data processing occurs to align the data into a common frame of reference (e.g. spatial or time).
- ii Data Association - Relationships among data points are identified. For example, in surveillance systems, a data association function would attempt to group all the measurements from different platforms (that is, any object that is carrying a sensor) that are associated with the same target. In an MCM/PHM system, oil, temperature and vibration measurements could be associated to the component they are measuring.
- iii State Estimation - In this step, the target's state is calculated from the measurements obtained from the previous levels.
- iv Identification - In the identification step, an attempt is made to predict the identity or classification of an object.

Level 2. Situation Assessment (SA) - The results from the previous level are interpreted to establish a relationship between the reconstructed entity and an observed event [8].

Level 3. Threat Assessment - In the threat assessment stage, the outcomes of different plans to remedy the situation are analyzed and the best course of action is predicted. This is typically a prediction function.

Level 4. Process Assessment - The process assessment phase is a global level in which the effectiveness and performance of the overall process (both hardware and software) are reviewed to identify possible means of improving the system. This phase in the JDL process involves planning and control.

III. Framework

In this section, we will relate the JDL data fusion model to the OSA-CBM model for MCM/PHM systems in ocean turbines.

Knowing where and when fusion should be performed is considered a primary issue of data fusion [11]. Our fusion model, shown in Figure 2, addresses this issue. This model combines the data fusion process of the JDL model with the layers in the OSA-CBM system architecture to produce an overall design model. By identifying at what point in the OSA-CBM architecture data fusion is occurring, we are also addressing data fusion as a cross cutting concern.

The proposed model was adopted from the JDL model which has been slightly modified to include both a Detection phase and an Action phase (as shown in the diagram). The Detection phase replaces the State Estimation and Identification steps in the JDL Model. Since there is no clear cut distinction between these two sub-steps of the Object Refinement phase for MCM/PHM systems, these steps have been combined to form the Detection phase. In this Detection step, any anomalies in data will be identified and associated with the component which generated that fault. A new phase — Action — has been added to correspond to Advisory Generation. In this phase, data from all the other phases will be combined to determine if and how the machine can self-adjust to optimize its current state.

The Process Assessment phase remains at a global level which can be accessed from any other phase in the data fusion model. This phase will need to be able to trigger adjustment of the data fusion process to ensure optimal performance while satisfying system requirements (which include minimizing the false positive rate). One researcher pointed out the inherent difficulty in evaluating the efficacy of the data fusion process in terms of the intended objectives [11]. There is, therefore, a need for a viable evaluation technique that would be used within the Process Assessment phase.

IV. Case Study: Ocean Turbine MCM/PHM

In this section, we will apply our model to an MCM/PHM system for ocean turbines. Unlike other machines, ocean turbines are intolerant of false positives due to the high costs of equipment retrieval. Therefore, an optimal data fusion approach must minimize the occurrence of false positives. Additionally, the data fusion system must be capable of handling real time data streaming from multiple, possibly heterogeneous, sensors. Fault detection must occur within a reasonable timeframe from the occurrence to minimize damage to the turbine.

As discussed in Section II, data fusion can occur at many different levels in the MCM/PHM architecture. Determining the type of data fusion algorithm suitable for each fusion level requires an analysis of the type of sensor data available as well as the type of inference that is desired. In other words, an algorithm that would output the system state as either "operational" or "malfunctioning" would differ from one needed to coordinate and integrate two waveforms.

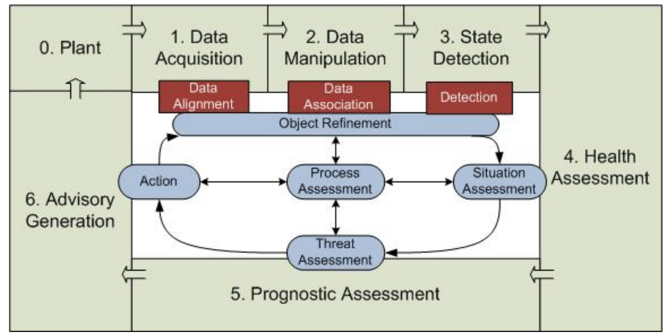

Figure 2. Data Fusion and MCM Model

While the model of Figure 2 indicates when data fusion is occurring, it provides no notion of how fusion will occur. Zooming down to the data fusion processes which are occurring at the different levels in the MCM/PHM architecture, we identified 3 unique types of data fusion that can occur within this architecture. These are labeled F1, F2 and F3 in Figures 3 and 4, and will be discussed in Sections IV-A, IV-B and IV-C respectively. From discussion in this section, it is clear that a single fusion process will occur across multiple levels of the hybrid framework proposed in Section 3, thus establishing data fusion as a cross cutting concern.

Before these sensor signals can be fused, we need to ensure that both signals originate from the same time interval. Since incoming data items are arriving continuously and consecutively, this flow of timed items is called a continuous timed data stream [12]. In future work, we will attempt to coordinate these streams through implementation of barrier synchronization [13], a well known coordination paradigm used in networking and in distributed and parallel computing.

A. Inter-Sensor Data Fusion

The first type of fusion, which occurs at boxes F1 in Figure 3, works closest to the raw data emanating from the sensors. At the lowest level, sensor data from two heterogeneous sources will be combined to validate signals, create features, and generate a unified output. In our context, we will be combining vibration signals from the accelerometer through channel C1 with velocity data from the tachometer through channel C3 to produce time synchronous average vibration features [14] through channel C1.3. Inter-sensor data fusion would occur during the Data Alignment phase of our hybrid framework (Figure 2) where the common frame of reference would be the tachometer signal.

In the inter-sensor data fusion process, the raw analogue signal that is received from the sensors must first be digitized through time into a sequence of digital values via a two step process using an analogue-to-digital converter (ADC). Sampling, the first step of the digitization process, is necessary to produce a signal in terms of discrete time and continuous values, while quantization, the second step, maps the set of continuous data values to a discrete set.

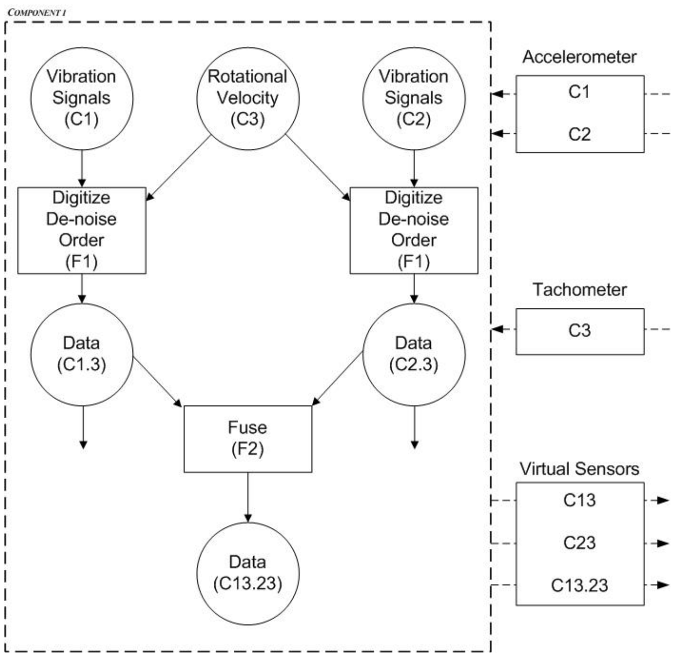

Figure 3. Intra-component fusion diagram

Once the data has been digitized, a low pass filter could be applied to de-noise the signal and improve the signal-to-noise ratio. Other de-noising techniques include a statistical multi-channel filtering approach [15], wavelet transforms, Wiener filter [16] and linear filters.

After undergoing a de-noising process, the vibration signal will then need to be normalized with respect to the tachometer through channel C3 in a process known as *ordering* or *order tracking*. By normalizing the accelerometer data with respect to the tachometer signal, we ensure that time samples are aligned to the same angular position of the component (e.g. a gear or shaft) to preserve the phase relationship. Additionally, since the rotational speed is not constant, order tracking is needed to distinguish non-stationary vibration data from transient vibrations caused by anomalies.

Time Synchronous Averaging (TSA) is a signal processing technique often used to perform inter-sensor data fusion and reduce noise. In TSA, interpolation is used to introduce new data points to the time series. The data is then divided into equal-sized blocks related to the synchronous signal being provided by channel C3, and these blocks are averaged together. The current block would start on the leading edge of the tachometer signal and end on the corresponding point preceding the following tachometer pulse. By taking sufficient averages, an improved estimate of the desired signal can be determined since any random noise will be eliminated [17].

On the opposite side of the component, a second accelerometer measuring the vibration of the same component but situated at a different angular position produces signal through channel C2. The entire fusion process F1 will be mirrored there to normalize the tachometer signal through channel C3 with respect to the accelerometer signal through channel C2 to generate signals through channel C2.3.

B. Intra-Component Data Fusion

In the second type of data fusion, denoted by F2, homogeneous data from multiple sources (channels C1.3 and C2.3) will be integrated to produce more accurate data than could have otherwise been generated from a single sensor. This may be between sensors measuring the same object but at different angular positions or between redundant sensors measuring the same object at the same angular position. In the former case, data from both sensors will need to be included in the final result since each sensor provides a unique perspective of the object. In the latter case, the data will help to identify sensor malfunction and would provide a mechanism for sensor validation. In either case, the fusion process F2 will output data through channel C13.23 in Figure 3.

Intra-component sensor fusion can be accomplished by any of three different levels of abstraction: signal/data-level fusion, feature-level fusion, or decision-level fusion. At the data-level, the data fusion algorithm is executed on the raw signals. Greater accuracy is obtained from doing so but this approach is only viable when the signals are of the same type (as would be through channels C1.3 and C2.3). At the feature level, feature extraction is done to produce a feature vector on each signal and the extracted feature vectors are fused. A sample feature set would be 320 bins of vibration data that correspond to different frequency ranges in a spectrogram. At the highest level of abstraction (decision-level fusion), each sensor processes their signals independently to produce a local estimate. These local estimates are then combined via a fusion process. By performing decision level fusion, features can be added or removed from the system without having to change the method of analysis. While this does provide a significant advantage, accuracy is sacrificed. Data level fusion will be performed here to maximize accuracy.

The concept of building an ensemble of classifiers to perform distributed classification has been adopted and successfully implemented over the past few years. One such application of ensembles for distributed classification was developed for node classification in peer-to-peer networks [18].

Folino, Pizzuti and Spezzano proposed in [19] an adaptive distributed ensemble classifier called StreamGP. In StreamGP, the local models are decision trees generated using Genetic Programming (GP), each containing its own local streaming data. In their approach, each node in the network was assumed to be continuously receiving blocks of data (labeled and unlabeled) over time in batches. The learning system would use the labeled data to train the ensemble and to update it when changes in the data are observed.

Such a dynamic approach could prove useful for this case study since the model evolves to deal with a new attack or fault. Consider, for example, a deployed ocean turbine in operation with a non-critical fault. If the turbine's MCM/PHM system detects a new fault, this information could be fed back to the Process Assessment phase which would initiate re-training of the ensemble classifier to generate an updated model. Using StreamGP to perform intra-component fusion will be analyzed in greater depth to ensure better performance and a lower false positive rate than a static ensemble.

C. Inter-Component Data Fusion

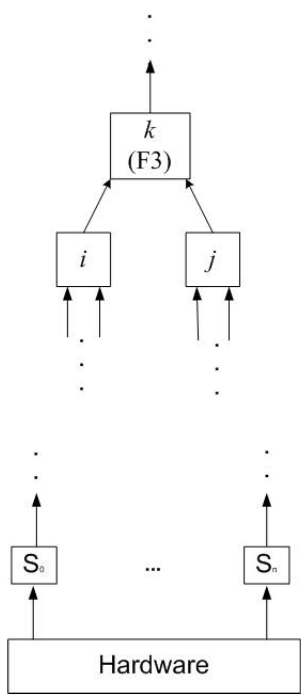

Figure 4. Inter-component fusion diagram

The health of a complex system cannot truly be determined from individual models and data analysis techniques. So, at our final level, we combine data from multiple components to enable system-level diagnosis and true prognostics. This level of fusion, referred to as knowledge fusion, involves integrating higher-order, possibly competing data and will be observed only in upper level nodes in the system architecture. Figure 4 diagrammatically represents this concept. Note that fusion types F1 and F2 will only occur within components i and j if they are lower level components (meaning that the input to these components are raw sensor signals), but F3 would occur within any other component to fuse the virtual sensor signals from any two components below it in the architecture. Inter-component fusion would occur during the Situation Assessment phase of Figures 1 and 2 as it establishes how a fault in a single component will affect the health and life expectancy of the entire system.

We now need to consider how to not only combine heterogeneous data from multiple sources, but how the fusion process will operate in the presence of discordance. Early research done in this area resulted in the formal definition of four possible ways to integrate input from these different types of components [20]. The four modes, known as Bower's Taxonomy of Fusion, are:

1) *complete unity*, wherein the data from all the components are combined without any consideration for the possible discord and hence no system is in place for detecting any disagreements;
2) *unity with awareness of discordance and the possibility of recalibration*, in which the conflicts are detectable and are reconciled by recalibrating the offending sensors;
3) *unity with awareness of discordance and tendency towards suppression*, in which disagreements are detectable but the information gathered from the offending sensors is temporarily ignored;
4) *no unity at all*

In complex environments, the mode of integration may rely on the situation so it is difficult (or may be impossible) to apply the same mode of integration throughout the entire system, since the system will need to adapt to changing environmental and operational conditions [21].

Because of the ability to incorporate human knowledge in determining the precedence of signals in the event of the conflict, an expert system could be built to distinguish the appropriate mode for a specific situation and perform inter-component fusion. Liu ands Liu previously implemented an expert system via a fuzzy group multiple attribute decision making method to perform this level of fusion in a machine condition monitoring system [22]. In their approach, the expert system was comprised of four modules, namely the diagnosis tree, a fuzzy group multiple attribute decision maker, a knowledge base in which each rule was associated with a confidence factor representing the level of uncertainty in the validity of the rule, and an inference engine.

The use of expert systems to accomplish this level of fusion will be investigated and adapted to the problem at hand.

V. Conclusions and Future Work

This paper defines an approach to data fusion for MCM/PHM systems in inaccessible, autonomous ocean systems. The proposed data fusion approach has been partitioned into three levels: inter–sensor or fusion of data between heterogeneous sensors, intra–component or data fusion between homogeneous sensors, and inter–component fusion which is performed at higher levels of the architecture to provide an overview of the health of the system.

At the inter–component and intra–component levels of fusion, the performance and false positive rates of distributed classification and expert systems must be carefully measured to determine the effectiveness of these techniques to our purposes. Additionally, further research will reveal a best way to represent and determine efficient inference rules for the expert system.

A synchronization technique for coordinating timed data streams generated from the various channels is necessary and essential. A more detailed specification and analysis of a barrier synchronization method for coordinating these data streams will be discussed in future work.

Future research will also involve formalizing a mathematical data fusion model for MCM/PHM systems for unmanned oceanic systems. To ensure that our approach is mathematically sound, a means of measuring the effectiveness of our data fusion framework will be defined.

Acknowledgements

The work discussed here grew from collaborations within the Prognostics and Health Monitoring (PHM) working group of the Center for Ocean Energy Technology (COET) at Florida Atlantic University and was funded through COET by the State of Florida.

References

[1] P.-P. Beaujean, T. M. Khoshgoftaar, J. C. Sloan, N. Xiros, and D. Vendittis, "Monitoring ocean turbines: a reliability assessment," in *Proceedings of the 15th ISSAT International Reliability and Quality in Design Conference*, 2009, pp. 367–371.

[2] J. Ye, K. Chen, T. Wu, J. Li, Z. Zhao, R. Patel, M. Bae, R. Janardan, H. Liu, G. Alexander, and E. Reiman, "Heterogeneous data fusion for alzheimer's disease study," in *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, (New York, NY, USA), ACM, 2008, pp. 1025–1033.

[3] D. Smith and S. Singh, "Approaches to multisensor data fusion in target tracking: A survey," *IEEE Trans. on Knowl. and Data Eng.*, vol. 18, no. 12, 2006, pp. 1696–1710.

[4] D. C. Conner, P. R. Kedrowski, and C. F. Reinholtz, "Multiple camera, laser rangefinder, and encoder data fusion for navigation of a differentially steered 3-wheeled autonomous vehicle," in *Proceedings of SPIE, the International Society for Optical Engineering*, 2001, pp. 76–83.

[5] M. Mjit, "Methodology for fault detection and diagnostics in an ocean turbine using vibration analysis and modeling," masters, Florida Atlantic University, 777 Glades Road, Boca Raton FL 33431, 2009.

[6] MIMOSA, "Osa-cbm v3.3.0," 2010.

[7] D. Raheja, J. Llinas, R. Nagi, and C. Romanowski, "Data fusion/data mining-based architecture for condition-based maintenance," *International Journal of Production Research*, vol. 44, jul 2006, pp. 2869–2887.

[8] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. Bryanston-Cross, "A review of data fusion models and architectures: towards engineering guidelines," *Neural Comput. Appl.*, vol. 14, no. 4, 2005, pp. 273–281.

[9] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the jdl data fusion model," *Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE*, vol. 3719, no. 1, 1999, pp. 430–441.

[10] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, and F. White, "Revisiting the jdl data fusion model ii," in *In P. Svensson and J. Schubert (Eds.), Proceedings of the Seventh International Conference on Information Fusion (FUSION 2004*, 2004, pp. 1218–1230.

[11] D. L. Hall and S. A. H. Mcmullen, *Mathematical Techniques in Multisensor Data Fusion (Artech House Information Warfare Library)*. March 2004.

[12] R. Herrtwich, "Timed data streams in continuous-media systems," Tech. Rep. TR-90-017, International Computer Science Institute, Berkley CA, May 1990.

[13] L. G. Valiant, "A bridging model for parallel computation," *Commun. ACM*, vol. 33, no. 8, 1990, pp. 103–111.

[14] R. Orsagh, J. Sheldon, and C. Klenke, "Prognostics/diagnostics for gas turbine engine bearings," in *Aerospace Conference, 2003. Proceedings. 2003 IEEE*, vol. 7, 8-15, 2003, pp. 3095 – 3103.

[15] A. Rao and D. Jones, "A denoising approach to multisensor signal estimation," *Signal Processing, IEEE Transactions on*, vol. 48, may 2000, pp. 1225 –1234.

[16] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. 1964.

[17] L. M, M. K, C. R, B. C, and M. K., "Review of vibration analysis methods for gearbox diagnostics and prognostics," *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology*, May 2000, pp. 623–634.

[18] P. Luo, H. Xiong, K. L"u, and Z. Shi, "Distributed classification in peer-to-peer networks," in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, (New York, NY, USA), ACM, 2007, pp. 968–976.

[19] G. Folino, C. Pizzuti, and G. Spezzano, "An adaptive distributed ensemble approach to mine concept-drifting data streams," in *ICTAI '07: Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence*, (Washington, DC, USA), IEEE Computer Society, 2007, pp. 183–188.

[20] C. Hiransoog and C. Malcolm, "Multi-sensor/knowledge fusion," in *Multisensor Fusion and Integration for Intelligent Systems, 1999. MFI '99. Proceedings. 1999 IEEE/SICE/RSJ International Conference on*, 1999, pp. 117 –122.

[21] J. Triesch and C. V. D. Malsburg, "Democratic integration: Self-organized integration of adaptive cues," *Neural Computation*, vol. 13, no. 9, 2001, pp. 2049–2074.

[22] S. Liu and S. Liu, "An efficient expert system for machine fault diagnosis.," *International Journal of Advanced Manufacturing Technology*, vol. 21, no. 9, 2003, pp. 691 – 698.